Conventions, Accuracy Metrics, Classification, Regression

Nipun Batra

July 19, 2025

IIT Gandhinagar

Demo

- Complete PoseNet Demo

Demo

- Complete PoseNet Demo
- Blog post from Google

Demo

- Complete PoseNet Demo
- Blog post from Google
- Rock Paper Scissors

Revision: What is Machine Learning

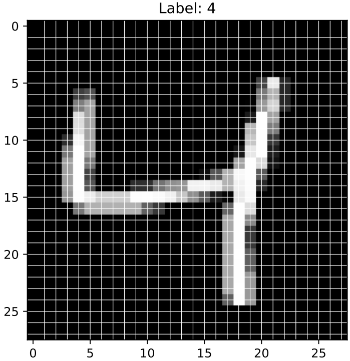
“Field of study that gives computers the ability to learn without being explicitly programmed” - Arthur Samuel [1959]

Revision: What is Machine Learning

"Field of study that gives computers the ability to learn without being explicitly programmed" - Arthur Samuel [1959]

Let us work on the digit recognition problem.

Notebook: rule-based-vs-ml.html



Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: $| + — + | +$ another vertically down $|$

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: $| + — + | +$ another vertically down |

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |
- The heights of each of the | need to be similar within tolerance

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |
- The heights of each of the | need to be similar within tolerance

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: $| + — + | +$ another vertically down $|$
- The heights of each of the $|$ need to be similar within tolerance
- Each of the $|$ can be slightly slanted. Similarly the horizontal line can be slanted.

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: $| + — + | +$ another vertically down |
- The heights of each of the | need to be similar within tolerance
- Each of the | can be slightly slanted. Similarly the horizontal line can be slanted.

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: $| + — + | +$ another vertically down $|$
- The heights of each of the $|$ need to be similar within tolerance
- Each of the $|$ can be slightly slanted. Similarly the horizontal line can be slanted.
- There can be some cases of 4 where the first $|$ is at 45 degrees

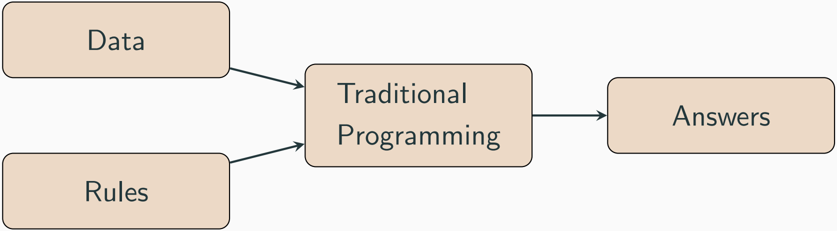
Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |
- The heights of each of the | need to be similar within tolerance
- Each of the | can be slightly slanted. Similarly the horizontal line can be slanted.
- There can be some cases of 4 where the first | is at 45 degrees

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: $| + — + | +$ another vertically down |
- The heights of each of the | need to be similar within tolerance
- Each of the | can be slightly slanted. Similarly the horizontal line can be slanted.
- There can be some cases of 4 where the first | is at 45 degrees
- There can be some cases of 4 where the width of each stroke is different

Traditional Programming vs Machine Learning



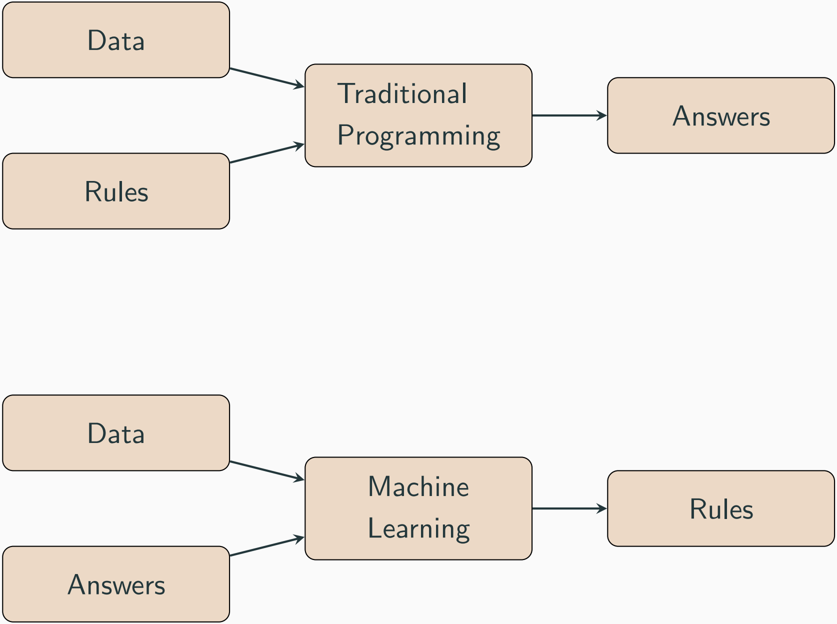
Data

Rules

Traditional Programming

Answers

Traditional Programming vs Machine Learning



Revision: What is Machine Learning

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.” - Tom Mitchell

First ML Task: Grocery store tomato quality prediction

Problem statement: You want to predict the quality of a tomato given its visual features.

Dataset

Imagine you have some past data on quality of tomatoes. What visual features do you think will be useful?

Dataset

Imagine you have some past data on quality of tomatoes. What visual features do you think will be useful?

Dataset

Imagine you have some past data on quality of tomatoes. What visual features do you think will be useful?

- Size

Dataset

Imagine you have some past data on quality of tomatoes. What visual features do you think will be useful?

- Size

Dataset

Imagine you have some past data on quality of tomatoes. What visual features do you think will be useful?

- Size
- Colour

Dataset

Imagine you have some past data on quality of tomatoes. What visual features do you think will be useful?

- Size
- Colour

Dataset

Imagine you have some past data on quality of tomatoes. What visual features do you think will be useful?

- Size
- Colour
- Texture

Sample Dataset

Here is our example dataset with tomato features:

Sample	**Colour**	**Size**	**Texture**	**Condition**
1	Orange	Small	Smooth	Good
2	Red	Small	Rough	Good
3	Orange	Medium	Smooth	Bad
4	Yellow	Large	Smooth	Bad

Useful Features

Is the sample number a useful feature for predicting quality of a tomato?

Useful Features

Is the sample number a useful feature for predicting quality of a tomato?

Answer: Usually no! Sample numbers are typically arbitrary identifiers and not meaningful features. Let us remove it.

Useful Features

Is the sample number a useful feature for predicting quality of a tomato?

Answer: Usually no! Sample numbers are typically arbitrary identifiers and not meaningful features. Let us remove it.

Let us modify our data table for now.

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

The training set consists of two parts:

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

The training set consists of two parts:

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

The training set consists of two parts:

1. Features (Input Variables)

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

The training set consists of two parts:

1. Features (Input Variables)

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

The training set consists of two parts:

1. Features (Input Variables)
2. Output or Response Variable

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathbb{R}^{n \times d}$) containing data of n samples each of which is d dimensional.

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathbb{R}^{n \times d}$) containing data of n samples each of which is d dimensional.

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathbb{R}^{n \times d}$) containing data of n samples each of which is d dimensional.
    - Thus, $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathbb{R}^{n \times d}$) containing data of n samples each of which is d dimensional.
   - Thus, $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathbb{R}^{n \times d}$) containing data of n samples each of which is d dimensional.
   - Thus, $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$
   - Example (after encoding): $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ (Orange=1, Small=0, Smooth=1)

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathbb{R}^{n \times d}$) containing data of n samples each of which is d dimensional.
   - Thus, $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$
   - Example (after encoding): $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ (Orange=1, Small=0, Smooth=1)

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathbb{R}^{n \times d}$) containing data of n samples each of which is d dimensional.
   - Thus, $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$
   - Example (after encoding): $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ (Orange=1, Small=0, Smooth=1)

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathbb{R}^{n \times d}$) containing data of n samples each of which is d dimensional.
   - Thus, $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$
   - Example (after encoding): $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ (Orange=1, Small=0, Smooth=1)

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathbb{R}^{n \times d}$) containing data of n samples each of which is d dimensional.
   - Thus, $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$
   - Example (after encoding): $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ (Orange=1, Small=0, Smooth=1)

Prediction Task

Estimate condition for unseen tomatoes (#5, 6) based on data set.

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad
Red	Large	Rough	?
Orange	Large	Rough	?

Testing Set

Testing set is similar to training set, but, does not contain labels for output variable.

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad
Red	Large	Rough	?
Orange	Large	Rough	?

Prediction Task

We hope to:

Prediction Task

We hope to:

Prediction Task

We hope to:

1. Learn f: Condition $= f$(colour, size, texture)

Prediction Task

We hope to:

1. Learn f: Condition $= f(\text{colour, size, texture})$

Prediction Task

We hope to:

1. Learn f: Condition $= f$(colour, size, texture)
2. From Training Dataset

Prediction Task

We hope to:

1. Learn f: Condition $= f$(colour, size, texture)
2. From Training Dataset

Prediction Task

We hope to:

1. Learn f: Condition $= f$(colour, size, texture)
2. From Training Dataset
3. To Predict the condition for the Testing set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad
Red	Large	Rough	?
Orange	Large	Rough	?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!

Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!

Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!
- Ideally - we want to predict “well” on all possible inputs. But, can we test that?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!
- Ideally - we want to predict “well” on all possible inputs. But, can we test that?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!
- Ideally - we want to predict “well” on all possible inputs. But, can we test that?
- No! Since, the test set is only a sample from all possible inputs.

Generalisation

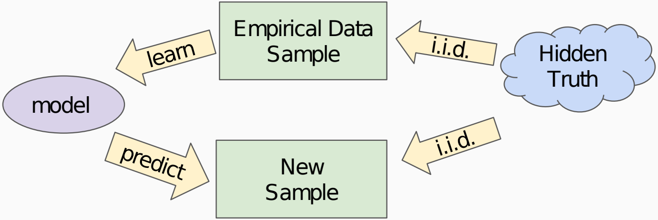


Image courtesy Google ML crash course

Generalisation

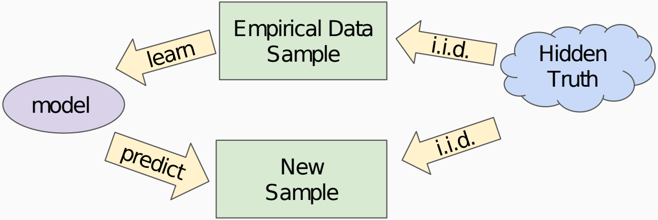


Image courtesy Google ML crash course

Both the training set and the test set are samples drawn from the hidden true distribution (also sometimes called population)

Generalisation

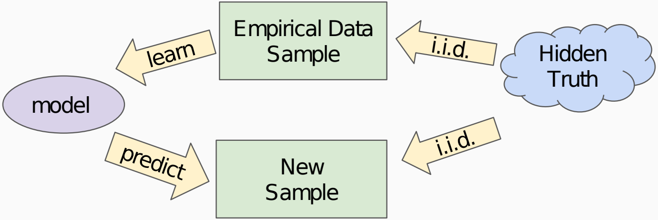


Image courtesy Google ML crash course

Both the training set and the test set are samples drawn from the hidden true distribution (also sometimes called population)

More discussion later once we study bias and variance

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?

Answer:

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?

Answer:

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?

Answer:

- # People (More people $\implies$ More Energy)

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?

Answer:

- # People (More people $\implies$ More Energy)

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?

Answer:

- # People (More people $\implies$ More Energy)
- Temperature (Higher Temp. $\implies$ Higher Energy)

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?

Answer:

- # People (More people $\implies$ More Energy)
- Temperature (Higher Temp. $\implies$ Higher Energy)

# People	**Temp (C)**	**Energy (kWh)**
4000	30	30
4200	30	32
4200	35	40
3000	20	?
1000	45	?

Classification vs Regression

- Classification

Classification vs Regression

- Classification

Classification vs Regression

- Classification

Classification vs Regression

- Classification
    - Output variable is discrete

Classification vs Regression

- Classification
    - Output variable is discrete

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes

Classification vs Regression

- Classification
    - Output variable is discrete
    - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \dots, k\}$ where k is number of classes
  - Examples - Predicting:

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)

Classification vs Regression

- Classification
    - Output variable is discrete
    - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
    - Examples - Predicting:
        - Will I get a loan? (Yes, No)

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression

Classification vs Regression

- Classification
    - Output variable is discrete
    - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
    - Examples - Predicting:
        - Will I get a loan? (Yes, No)
        - What is the quality of fruit? (Good, Bad)
- Regression

Classification vs Regression

- Classification
    - Output variable is discrete
    - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
    - Examples - Predicting:
        - Will I get a loan? (Yes, No)
        - What is the quality of fruit? (Good, Bad)
- Regression
    - Output variable is continuous

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathbb{R}$

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathbb{R}$

Classification vs Regression

- Classification
    - Output variable is discrete
    - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
    - Examples - Predicting:
        - Will I get a loan? (Yes, No)
        - What is the quality of fruit? (Good, Bad)
- Regression
    - Output variable is continuous
    - i.e. $y_i \in \mathbb{R}$
    - Examples - Predicting:

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathbb{R}$
  - Examples - Predicting:

Classification vs Regression

- Classification
    - Output variable is discrete
    - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
    - Examples - Predicting:
        - Will I get a loan? (Yes, No)
        - What is the quality of fruit? (Good, Bad)
- Regression
    - Output variable is continuous
    - i.e. $y_i \in \mathbb{R}$
    - Examples - Predicting:
        - How much energy will campus consume?

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathbb{R}$
  - Examples - Predicting:
    - How much energy will campus consume?

Classification vs Regression

- Classification
    - Output variable is discrete
    - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
    - Examples - Predicting:
        - Will I get a loan? (Yes, No)
        - What is the quality of fruit? (Good, Bad)
- Regression
    - Output variable is continuous
    - i.e. $y_i \in \mathbb{R}$
    - Examples - Predicting:
        - How much energy will campus consume?
        - How much rainfall will fall?

Metrics for Classification

$$\overset{\text{Prediction } (\hat{y})}{\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix}} \qquad \overset{\text{Ground Truth } (y)}{\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Bad} \end{pmatrix}}$$

Ground Truth: From the actual training set
Prediction: Made by the model

Accuracy

$$\begin{array}{ccc} & \text{Prediction } (\hat{y}) & \text{Ground Truth } (y) \\ \begin{array}{c} \checkmark \\ \checkmark \\ \\ \\ \checkmark \end{array} & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix} & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Bad} \end{pmatrix} \end{array}$$

Accuracy

$$\begin{array}{c} \text{Prediction } (\hat{y}) \\ \begin{array}{c}\checkmark\\ \checkmark\\ \\ \\ \checkmark\end{array}\begin{pmatrix}\text{Good}\\ \text{Good}\\ \text{Good}\\ \text{Good}\\ \text{Bad}\end{pmatrix}\end{array} \qquad \begin{array}{c} \text{Ground Truth } (y) \\ \begin{pmatrix}\text{Good}\\ \text{Good}\\ \text{Bad}\\ \text{Bad}\\ \text{Bad}\end{pmatrix}\end{array}$$

$$\begin{aligned} \text{Accuracy} &= \frac{|\{i : y_i = \hat{y}_i\}|}{n} \\ &= \frac{3}{5} = 0.6 \end{aligned}$$

Mathematical Notation: Set Cardinality and Indicator Functions

Mathematical Notation: Set Cardinality and Indicator Functions

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
    - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: “Number of indices i such that $y_i = \hat{y}_i$”
  - Counts how many samples satisfy the condition

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"
  - Counts how many samples satisfy the condition

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
    - Reads as: “Number of indices i such that $y_i = \hat{y}_i$”
    - Counts how many samples satisfy the condition

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"
  - Counts how many samples satisfy the condition
- **Alternative: Indicator function notation**

$$\text{Accuracy} = \frac{\sum_{i=1}^{n} \mathbf{1}[y_i = \hat{y}_i]}{n}$$

$$\text{where } \mathbf{1}[\text{condition}] = \begin{cases} 1 & \text{if condition is true} \\ 0 & \text{if condition is false} \end{cases}$$

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"
  - Counts how many samples satisfy the condition
- **Alternative: Indicator function notation**

$$\text{Accuracy} = \frac{\sum_{i=1}^{n} \mathbf{1}[y_i = \hat{y}_i]}{n}$$

$$\text{where } \mathbf{1}[\text{condition}] = \begin{cases} 1 & \text{if condition is true} \\ 0 & \text{if condition is false} \end{cases}$$

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"
  - Counts how many samples satisfy the condition
- **Alternative: Indicator function notation**

$$\text{Accuracy} = \frac{\sum_{i=1}^{n} \mathbf{1}[y_i = \hat{y}_i]}{n}$$

$$\text{where } \mathbf{1}[\text{condition}] = \begin{cases} 1 & \text{if condition is true} \\ 0 & \text{if condition is false} \end{cases}$$

- Both notations are mathematically equivalent and commonly used in ML literature

Types of Data: Imbalanced Classes

$$\begin{matrix} \text{1 sample } \{ \\ \text{100 samples } \left\{ \vphantom{\begin{matrix} a\\ a\\ a\\ a \end{matrix}} \right. \end{matrix} \begin{pmatrix} \text{Bad} \\ \text{Good} \\ \text{Good} \\ \dots \\ \text{Good} \end{pmatrix} \qquad \text{Imbalanced Classes}$$

Types of Data: Imbalanced Classes

$$\begin{array}{r} 1\text{ sample}\ \{ \\ 100\text{ samples}\ \left\{\begin{array}{l} \\ \\ \\ \\ \end{array}\right. \end{array} \begin{pmatrix} \text{Bad} \\ \text{Good} \\ \text{Good} \\ \dots \\ \text{Good} \end{pmatrix} \qquad \text{Imbalanced Classes}$$

Cases for this:

- Cancer Screening

Types of Data: Imbalanced Classes

$$\begin{array}{r} \text{1 sample}\left\{\vphantom{\text{Bad}}\right. \\ \text{100 samples}\left\{\begin{array}{c}\text{Good}\\ \text{Good}\\ \dots\\ \text{Good}\end{array}\right. \end{array} \begin{pmatrix}\text{Bad}\\ \text{Good}\\ \text{Good}\\ \dots\\ \text{Good}\end{pmatrix} \qquad \text{Imbalanced Classes}$$

Cases for this:

- Cancer Screening
- Planet Detection

Accuracy Metrics: Precision

$$\begin{array}{rc} & \text{Prediction } (\hat{y}) \\ \begin{array}{r} \rightarrow \checkmark \\ \rightarrow \checkmark \\ \rightarrow \\ \rightarrow \\ \end{array} & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix} \end{array} \qquad \begin{array}{c} \text{Ground Truth } (y) \\ \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Good} \end{pmatrix} \end{array}$$

$$\text{Precision} = \frac{|\{i : y_i = \hat{y}_i = \text{Good}\}|}{|\{i : \hat{y}_i = \text{Good}\}|} = \frac{2}{4} = 0.5$$

“the fraction of relevant instances among the retrieved instances”, i.e. “out of the number of times we predict Good, how many times is the condition actually Good”

Accuracy Metrics: Precision

$$\begin{array}{r} \rightarrow \checkmark \\ \rightarrow \checkmark \\ \rightarrow \\ \rightarrow \\ \end{array} \overset{\text{Prediction } (\hat{y})}{\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix}} \qquad \qquad \overset{\text{Ground Truth } (y)}{\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Good} \end{pmatrix}}$$

$$\text{Precision} = \frac{|\{i : y_i = \hat{y}_i = \text{Good}\}|}{|\{i : \hat{y}_i = \text{Good}\}|} = \frac{2}{4} = 0.5$$

“the fraction of relevant instances among the retrieved instances”, i.e. “out of the number of times we predict Good, how many times is the condition actually Good”

Accuracy Metrics: Recall

$$\begin{array}{cc} & \text{Prediction } (\hat{y}) \\ \begin{array}{l} \rightarrow \checkmark \\ \rightarrow \checkmark \\ \\ \\ \rightarrow \end{array} & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix} \end{array} \qquad \begin{array}{c} \text{Ground Truth } (y) \\ \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Good} \end{pmatrix} \end{array}$$

$$\text{Recall} = \frac{|\{i : y_i = \hat{y}_i = \text{Good}\}|}{|\{i : y_i = \text{Good}\}|} = \frac{2}{3} = 0.67$$

“the fraction of the total amount of relevant instances that were actually retrieved”

Types of Data: Imbalanced Classes

Given predictions of whether a tissue is cancerous or not ($n = 100$).

$$\text{Prediction } (\hat{y}) \qquad\qquad \text{Ground Truth } (y)$$

$$\begin{array}{r}\rightarrow\\ \\ \\ \\ \\ \end{array}\begin{pmatrix} \text{Yes} \\ \text{No} \\ \text{No} \\ \dots \\ \text{No} \end{pmatrix} \qquad \begin{array}{r} \\ \\ \\ \\ \rightarrow\end{array}\begin{pmatrix} \text{No} \\ \text{No} \\ \dots \\ \text{No} \\ \text{Yes} \end{pmatrix}$$

Types of Data: Imbalanced Classes

Given predictions of whether a tissue is cancerous or not ($n = 100$).

$$\begin{array}{cc} & \text{Prediction } (\hat{y}) \\ \begin{array}{c} \rightarrow \\ \\ \\ \\ \end{array} & \begin{pmatrix} \text{Yes} \\ \text{No} \\ \text{No} \\ \dots \\ \text{No} \end{pmatrix} \end{array} \qquad \begin{array}{cc} & \text{Ground Truth } (y) \\ \begin{array}{c} \\ \\ \\ \\ \rightarrow \end{array} & \begin{pmatrix} \text{No} \\ \text{No} \\ \dots \\ \text{No} \\ \text{Yes} \end{pmatrix} \end{array}$$

$$\text{Accuracy} = \frac{98}{100} = 0.98$$

$$\text{Recall} = \frac{0}{1} = 0$$

$$\text{Precision} = \frac{0}{1} = 0$$

Accuracy Metrics: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	0	1
	No	1	98

Accuracy Metrics: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	0	1
	No	1	98

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

Accuracy Metric: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$$

Accuracy Metric: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$$

Accuracy Metric: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$$

Accuracy Metric: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$$

Accuracy Metrics: F-Score

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

$$\mathrm{F}- \text{Score} = \frac{2\times \text{ Precision } \times \text{ Recall}}{\text{Precision } + \text{ Recall}}$$

Accuracy Metrics: Matthew's Correlation Coefficient

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

$$\text{Matthew's correlation coefficient} = \frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

Accuracy Metrics: Example

For the data given below, calculate:

$$\begin{array}{lcc} & \text{G.T. Positive} & \text{G.T. Negative} \\ \text{Pred Positive} & 90 & 4 \\ \text{Pred Negative} & 1 & 1 \end{array}$$

Precision $=$?

Recall $=$?

F-Score $=$?

Matthew's Coeff. $=$?

Accuracy Metrics: Answer

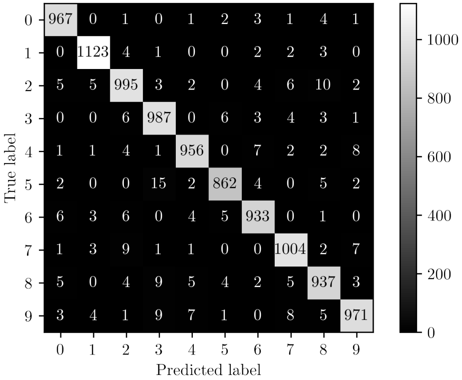
For the same data

$$\begin{array}{rc} & \begin{array}{cc} \text{G.T. Positive} & \text{G.T. Negative} \end{array} \\ \begin{array}{r} \text{Pred Positive} \\ \text{Pred Positive} \end{array} & \begin{pmatrix} \quad 90 \quad & \qquad 4 \qquad \\ \quad 1 \quad & \qquad 1 \qquad \end{pmatrix} \end{array}$$

Precision $= \frac{90}{94}$
Recall $= \frac{90}{91}$
F-Score $= 0.9524$
Matthew's Coeff. $= 0.14$

Confusion Matrix for multi-class classification

Notebook: confusion-mnist.html



True label \ Predicted label	0	1	2	3	4	5	6	7	8	9
0	967	0	1	0	1	2	3	1	4	1
1	0	1123	4	1	0	0	2	2	3	0
2	5	5	995	3	2	0	4	6	10	2
3	0	0	6	987	0	6	3	4	3	1
4	1	1	4	1	956	0	7	2	2	8
5	2	0	0	15	2	862	4	0	5	2
6	6	3	6	0	4	5	933	0	1	0
7	1	3	9	1	1	0	0	1004	2	7
8	5	0	4	9	5	4	2	5	937	3
9	3	4	1	9	7	1	0	8	5	971

Metrics for Regression MSE & MAE

$$\underset{\text{Prediction } (\hat{y})}{\begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix}} \qquad \underset{\text{Ground Truth } (y)}{\begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix}}$$

$$\text{Mean Squared Error (MSE)} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}$$

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\text{MSE}}$$

Accuracy Metrics: MAE & ME

$$\begin{array}{cc} \text{Prediction } (\hat{y}) & \text{Ground Truth} \\ \begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix} & \begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix} \end{array}$$

$$\text{Mean Absolute Error (MAE)} = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}$$

$$\text{Mean Error} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n}$$

Accuracy Metrics: MAE & ME

$$\begin{array}{cc} \text{Prediction } (\hat{y}) & \text{Ground Truth} \\ \begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix} & \begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix} \end{array}$$

$$\text{Mean Absolute Error (MAE)} = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}$$

$$\text{Mean Error} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n}$$

Is there any downside with using mean error?

Accuracy Metrics: MAE & ME

$$\begin{array}{c} \text{Prediction } (\hat{y}) \\ \begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix} \end{array} \qquad \begin{array}{c} \text{Ground Truth} \\ \begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix} \end{array}$$

$$\text{Mean Absolute Error (MAE)} = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}$$

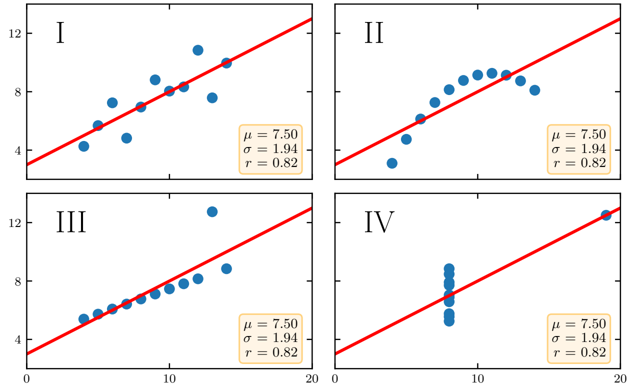
$$\text{Mean Error} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n}$$

Is there any downside with using mean error?

Errors can get cancelled out

The Importance of Plotting

Notebook: anscombe.html



Anscombe's Quartet

Dummy Baselines

Notebook: dummy-baselines.html

The Importance of Plotting

Property	Value	Across datasets
mean(X)	9	exact
mean(Y)	7.5	up to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	up to 2 decimal places